CentER

Discussion Paper

No. 2004–62

# AGGREGATION OF PRODUCTIVITY INDICES: THE ALLOCATIVE EFFICIENCY CORRECTION

By M.H. ten Raa

July 2004

ISSN 0924-7815

TILBURG UNIVERSITY

# Aggregation of Productivity Indices:
# The Allocative Efficiency Correction

Thijs ten Raa
Tilburg University
5000 LE Tilburg
the Netherlands
tenRaa@UvT.nl

February 26, 2004

**Abstract.** Industry productivity is obtained by aggregation of firm productivities and inclusion of the appropriate allocative efficiency terms, one for each firm. This paper identifies the latter correction terms.

**Keywords:** Aggregation, productivity, allocative efficiency

**JEL codes:** O47, C43

# 1. Introduction

Firms' productivity indices do not sum to the industry productivity index, except when production is linear in the sense that marginal rates of substitution and marginal rates of transformation are constant and these constants are common to the firms (Blackorby and Russell, 1999). The trouble is that industry productivity is influenced not only by the performance of firms, but also by the allocation of resources between the firms. In an attempt to salvage the aggregation of productivity, Färe and Primont (2003) show that if all firms are allocatively efficient and their technologies admit time-invariant quadratic approximations, then the productivity indices can be aggregated. Unfortunately, the Färe-Primont conditions are also prohibitively restrictive. To me the bottom line seems to be that the determination of industry productivity requires not only the aggregation of firm productivities, but also the inclusion of some allocative efficiency terms. This paper identifies those correction terms.

Unfortunately, the literature is loaded with formulas. Part of the blame can be put on the mix of conceptual and approximation issues. To keep the analysis transparent, I focus on concepts and side step approximation issues simply by working in continuous time. I also simplify the concept of a productivity indicator. Färe and Primont (2003) use the Luenberger indicator, which is based on the distance to the frontier along some direction in commodity space; they remain silent about the choice of direction. Now Woertman and ten Raa (2004) argue that for quasi-linear functions the direction is determined by the linear commodity component, as in that case Luenberger's measure is equal to both the compensating and the equivalent variations. This is in the context of consumtion theory, but the implication for production theory is that the appropriate Luenberger direction is along *output*, at least for single-output industries. This observation reduces the distance function to the output gap and, as we shall see, the derived productivity indicator to the Solow residual. It makes the analysis so crisp that the extension to multi-output industries becomes obvious.

## 2. Indices

I introduce the formalities. Single-output firm $k$ maps input vector $x_k(t)$ in output scalar $y_k(t) \leq F_k(x_k(t), t)$, where $F_k(\cdot, t)$ is its production function at time $t$. (Parameter $t$ shifts the production function, or what we call technical change.) $F'_k$ denotes the vector of marginal products (partial derivatives with respect to inputs, not time). As usual, a dot denotes a time derivative: $\dot{x}_k(t)$ is the time derivative of $x_k(t)$. Luenberger's output based distance function is given by

$$D_k(x_k(t), y_k(t), t) = F_k(x_k(t), t) - y_k(t) \tag{1}$$

and measures the *output gap*. In general, even without the quasi-linear structure of (1), the distance function measures *in*efficiency. *Efficiency change* is therefore defined by *minus* the change in the distance function:

$$EC_k = -(d/dt)D_k \tag{2}$$

The distance to the frontier may grow without any change in inputs or outputs, simply because the frontier shifts out. This is called *technical change*. It is defined by the partial derivative of the distance function with respect to time:

$$TC_k = (\partial/\partial t)D_k$$

(3)

The sum of efficiency change and technical change defines *productivity change*:

$$PC_k = EC_k + TC_k \tag{4}$$

Application of the chain rule to (2) and addition of (3) transforms (4) into

$$PC_k = -(\partial/\partial x_k)D_k \cdot \dot{x}_k(t) - (\partial/\partial y_k)D_k \cdot \dot{y}_k(t) \tag{5}$$

The discrete time approximation of (5) is what Chambers, Färe, and Grosskopf (1996) call the Luenberger productivity index. That index is the point of departure of Färe and Primont's (2003) aggregation analysis. In case of the quasi-linear structure of (1), expression (5) simplifies quite dramatically into

$$PC_k = \dot{y}_k(t) - F'_k(x_k(t), t)\cdot\dot{x}_k(t) \tag{6}$$

In other words, productivity change is equal to firm $k$'s *Solow residual* between its output change and input changes, where the latter are weighted by their marginal product values.

# 3. Aggregation

The more standard Solow residual is at the macro level, or, in the context of the present literature, the industry level. For this we need the *industry* distance function or output gap. Now potential output is determined by:

$$\max \sum F_k(\xi_k, t) \text{ subject to } \sum \xi_k = \sum x_k(t) \tag{7}$$

Since the optimal allocation, $(\xi_k)$, depends on time (through the constraint and the objective function), let me denote it by $(x_k^*(t))$. The crucial trouble behind the (negative) aggregation results of Blackorby and Russell (1999) and Färe and Primont (2003) is that attainment of the optimal industry output requires not only a push of the firms to their respective frontiers, from $y_k(t)$ to $F_k(\cdot, t)$, but also a reallocation of resources between them, that is from $x_k(t)$ to $x_k^*(t)$. The benefit of the latter reallocation is simply missed when firm efficiency indices are aggregated, without correction. The missing element is the potential allocative efficiency gain; it will be derived next.

As a first observation, notice that potential output, (7), is a function of *total input*, $x(t) = \sum x_k(t)$. Hence we may denote the solution to (7) by $F(x(t), t)$, and, therefore, the industry output gap, see (1), is

$$D(x(t), y(t), t) = F(x(t), t) - y(t) \tag{8}$$

where the last term is defined by $y(t) = \sum y_k(t)$. The productivity analysis of the firm can now be applied to the industry. In particular, (6) becomes

$$PC = \dot{y}(t) - F'(x(t), t) \cdot \dot{x}(t) \tag{9}$$

The "aggregation problem" consists of interrelating the micro- and macro-productivity changes, (6) and (9). This boils down to an analysis of the industry production function, $F$, which is the solution to (7). Now denote the Lagrange multipliers of the (vector) constraint in (7) by vector $w$. Since Lagrange multipliers measure the sensitivity of the objective function, $F$, with respect to the bounds in the constraints, $x(t)$, we have

$$w = F'(x(t), t) \tag{10}$$

Now the first order condition of (7) with respect to $\xi_k$ reads, in the optimum,

$$F_k'(x_k^*(t), t) = w \tag{11}$$

This is the well-known result that efficiency implies the equalization of marginal productivities. Substitution of (11) in (10) and subsequently in (9) yields

$$PC = \sum[\dot{y}_k(t) - F_k'(x_k^*(t), t) \cdot \dot{x}_k(t)] \tag{12}$$

Comparison of this result with (6) shows that *if* $F_k'(x_k^*(t), t) = F_k'(x_k(t), t)$, *then* aggregation is perfect, in the sense that $PC = \sum PC_k$. This condition is indeed fulfilled if marginal productivities are constant, an observation that confirms the result of Blackorby and Russell (1999). The condition is also fulfilled if the mixes of the observed input vectors are right, i.e. if the observed inputs $x_k(t)$ are collinear with the optimal ones $x_k^*(t)$, and returns to scale are constant, an observation that confirms the result of Färe and Primont (2003). If none these conditions are fulfilled, we must make a correction. In fact, the connection between (6) and (12) is:

$$PC = \sum PC_k + \sum[F_k'(x_k(t), t) - F_k'(x_k^*(t), t)]\cdot \dot{x}_k(t) \qquad (13)$$

It is interesting that the correction consists of a sum of terms, one for each firm. For each firm the correction measures the excess marginal productivities (over and above the competitive, economy-wide ones), weighted by the changes in inputs. The difference in brackets is the excess rate of return, or the difference between the private and social values of inputs.

It is not difficult to understand the correction expression. Suppose firm $k$ is underendowed with input 1. Then input 1 is relatively scarce at firm 1, hence will carry a high marginal product or supernormal private value. But the latter is used as a weight in the Solow residual of firm $k$, where the input change contributes *negatively*. In short, the scarcity of input 1 causes a downward bias in the Solow residual of firm $k$ when the private value weight is used instead of the social value. The positive correction term (the excess rate of return times the change in the input at firm $k$) offsets the bias.

The aggregation bias of productivity changes can go either way. In terms of efficiency *levels*, however, it goes one way, a fact that is exceedingly simple to demonstrate. The solution to (7) exceeds the value without reallocations:

$$F(x(t), t) \geq \sum F_k(x(t), t) \qquad (14)$$

In view of (1) and (8) it follows that

$$D(\sum x_k(t), \sum y(t)_k, t) \geq \sum D_k(x_k(t), y_k(t), t) \qquad (15)$$

Thus, industry inefficiency exceeds aggregate firms' inefficiency. The difference, of courses, is the allocative inefficiency.

# 4. Concluding remarks

Aggregate productivity is the sum of firm productivities *and* firm allocative efficiency changes. A firm's allocative efficiency change is measured by its excess marginal productivities (over and above the competitive economy wide ones), weighted by input changes.

We have derived this result for quasi-linear output gaps, $F(x(t),t) - y(t) \geq 0$. The extension to general production structures, $F(x(t), -y(t), t) \geq 0$, is obvious. If we redefine $(x(t), -y(t))$ as net input vector $x(t)$, we may drop $y_k(t)$ from productivity change (6) (and likewise for aggregate productivity change (9)), and the decomposition formula (13) remains valid.

Strictly speaking, the program defining the industry production function should feature nonnegativity constraints. However, the modification is a straightforward application of the Kuhn-Tucker conditions. The first order condition, (11), is replaced by $F_k'(x_k^*(t), t) \leq w$, with component slack only if the corresponding optimal input is zero. In formula (13) replacement of $F_k'(x_k^*(t), t)$ by $w$ takes care of this condition. A more elegant way to handle nonnegativity is the generalization to general production structures, $F(x(t), -y(t), t) \geq 0$, which accommodates it easily.

Not surprisingly, considering the theoretical state of affairs, the correction for allocative efficiency is overlooked in the applied econometric literature, such as Jorgenson, Ho, and Stiroh (2003). This paper fills the gap.

# References

Blackorby, C. and R.R. Russell, "Aggregation of Efficiency Indices," *Journal of Productivity Analysis* 12, 1, 5-20 (1999)

Färe, R. and D. Primont, "Luenberger Productivity Indicators: Aggregation Across Firms," *Journal of Productivity Analysis* 20, 3, 425-435 (2003)

Jorgenson, D., M. Ho, and K. Stiroh, "Growth of US Industries and Investments in Information Technology and Higher Education," *Economic Systems Research* 15, 3, 279-325 (2003)

Woertman, W. and Th. ten Raa, "Benefit and Expenditure Functions," draft, Tilburg University (2004)